知识管理论坛 ISSN 2095-5472 CN11-6036/C
Knowledge Management Forum E-mail: kmf@mail.las.ac.cn
http://www.kmf.ac.cn

【专稿】

网络舆情受众失范行为靶向引导的技术框架构建

黄微[1] 刘熠[1] 郭苏琳[1,2]
[1] 吉林大学管理学院 长春 130022
[2] 吉林建筑大学建筑与规划学院 长春 130119

摘要：[**目的 / 意义**] 网络舆情受众的失范行为是舆情出现剧烈变化的重要因素。本文重点研究失范行为靶向引导技术框架的构建，为失范行为实施引导提供技术手段。[**方法 / 过程**] 在网络舆情受众失范行为靶向引导分析的基础上，建立框架模型，并对框架模型的各个模块进行具体描述；梳理各模块功能实现所需要的核心技术和进一步研究的思路，并以微博舆情为例对框架进行了验证。[**结果 / 结论**] 靶向引导是理论与应用相结合的方法，网络舆情受众失范行为的靶向引导技术框架模型，能够从技术层面解决目前引导策略过于宏观失之具体的缺陷，增强了靶向引导的可操作性。

关键词：网络舆情 失范行为 靶向引导
分类号：G206
DOI：10.13266/j.issn.2095-5472.2020.015

引用格式：黄微 , 刘熠 , 郭苏琳 . 网络舆情受众失范行为靶向引导的技术框架构建 [J/OL]. 知识管理论坛 , 2020, 5(3): 159-174[引用日期]. http://www.kmf.ac.cn/p/210/.

1 前言

目前，自媒体的准入门槛不高，活跃度已经达到了前所未有的程度。由于缺乏“守门人”对信息内容和行为的把关，一部分网络舆情信息受众在参与舆情事件的过程中发生的信息行为不够规范，对网络生态环境造成了不良影响。因此，对网络舆情受众失范行为的发现与靶向引导成为当前研究的热门话题。

当前，网络舆情受众失范行为靶向引导技术的研究以高校引导和政府引导等多种方式并存。在高校引导方面，由于青年学子思想活跃，但未经历广泛的社会实践，因而具有行为冲动、易受他人影响、思维逻辑简单等特点 [1]。作为网络舆情受众中的重要群体，高校大学生又较易形成网络舆论 [2]。因此提高高校舆论引导力，稳固高校思想阵地的高校引导研究 [3] 是现阶段

基金项目：本文系国家社会科学基金重大项目“大数据驱动的社交网络舆情主题图谱构建及调控策略研究”（项目编号：18ZDA310）研究成果之一。
作者简介：黄微 (ORCID：0000-0003-0448-9563)，教授，博士，博士生导师，E-mail：huangwei@jlu.edu.cn；刘熠 (ORCID：0000-0002-7360-2091)，博士研究生；郭苏琳（ORCID:0000-0002-9941-4988)，博士研究生。
收稿日期：2020-03-30 发表日期：2020-06-24 本文责任编辑：刘远颖

舆情引导研究的重点。在政府引导方面，政府舆情引导策略更为复杂，舆情引导机制规制需要具有长效性、协调性和互动性，不仅需要政府、媒体和公众的同时作为，还要制定相应的法律法规 [4]。在政府舆情引导研究中，罗潇潇等人认为政府在处理突发公共事件时，需及时发布权威信息并建立畅通的信息沟通渠道，以正确控制和引导舆论 [5]。王晰巍等认为政府在引导舆情的过程中应当对移动端平台与非移动端平台用户之间的差异性予以重视 [6]。赵晓航等认为政府对舆情的回应须在危机发生初期及时发声，以防疑惑心理转化为负面情绪；同时应注重与舆论的互动，防止次生舆情的产生，以完成舆情引导工作 [7]。

现有网络舆情引导研究主要集中在宏观层面，多以逐条梳理网络舆情应对、治理、善后策略为主，引导方略在与实际接轨、开展实施方面略显单薄。在本文研究中，以网络舆情事件为背景，以研究网络舆情受众在不同事件中的失范行为为核心，针对不同失范行为，提出了靶向引导的技术框架，试图以信息技术为基础，增强更有效、更精准的技术型引导功能，从而弥补现阶段网络舆情引导操作性不强的问题，为网络舆情引导提供新的思路。

网络舆情受众失范行为靶向引导的理论基础，是基于托尔曼的 S-O-R（刺激 - 机体 - 反应）理论，也吸纳了较为成熟的情境理论 [8]、AVP 情感维度理论 [9]、曾盛泉的网络舆情应对理论 [10]、自然语言处理方法、语义识别方法等。通过对以上理论与方法的整合与利用，最终构建了网络舆情受众失范行为靶向引导的技术框架模型。

2 相关概念的界定

2.1 失范行为的概念界定

鱼震海认为，网络媒体的失范行为是违背社会规范不适当的使用互联网从而背离主流社会要求的行为偏差 [11]。谢晖认为，失范行为包括了舆论暴力、舆论绑架、舆论极端与舆论失真 [12]。杨祺认为，失范行为背离了社会主义核心价值观，会对社会造成不良影响 [13]。综合上述各类关于失范行为的定义，本文认为网络舆情受众失范行为是指所有在网络舆情事件中产生负面影响、引领舆情错误走向、导致舆情危机的网络舆情受众行为。

2.2 靶向引导的概念界定

面向网络舆情受众失范行为的靶向引导，是以网络舆情受众、网络舆情事件、网络舆情受众失范行为 3 个维度为靶向的，靶向引导是在 3 个维度层层递进，层层聚焦的基础上实施的。其分析过程是，通过分析网络舆情事件的背景，获取事件独立特征，判断具有独立特征的受众是否具备参与到网络舆情事件的潜质，而后根据受众的独立特征，对其参与该网络舆情事件的参与行为进行拟合预测，从而根据拟合预测结果提出具有针对性的靶向引导策略。靶向引导是以机器学习算法为基础的，自动根据各维度特征提出对应引导策略的技术方法，其显著特点是实现舆情处置准确化与自动化。

3 网络舆情受众失范行为靶向引导技术框架模型构建的目标与原则

3.1 网络舆情受众失范行为靶向引导技术框架模型构建的目标

面向网络舆情信息受众靶向引导的技术框架模型构建的重点是达成网络舆情信息受众拟合、事件拟合与受众行为预测从而实施靶向引导，包括实现网络舆情信息受众靶向引导的构成要素、实施流程与技术支持 3 个重要方面。构建面向网络舆情靶向引导的技术框架模型需要遵循数据实时获取、数据并行处理、靶向匹配、全自动输出 4 个基本步骤和目标流程。

（1）数据实时获取，即在整个框架模型下，任何模块的数据都在不断地更新与丰富，因此数据需要实时获取以达到靶向引导的时效性。

（2）数据并行处理，即在框架模型下，个体分型模块、事件匹配模块、行为匹配模块与触发点发现模块遵循分布式并行处理的要求。

知识管理论坛，2020(3):159-174
DOI：10.13266/j.issn.2095-5472.2020.015

同时受众画像、舆情事件感知，均按照多线程分布式计算，并不存在时间的先后顺序。

（3）靶向匹配，即该框架需要根据网络舆情信息受众、事件情境与网络舆情信息受众在事件情境下发生的失范行为，匹配出更合理的引导策略与引导路径，达到靶向引导的目的。

（4）自动输出，即整个系统在框架中的运行，不需要借助人工，从数据获取、数据处理到受众画像、舆情事件感知与失范行为拟合再到靶向引导是一个自动的过程。

3.2 网络舆情受众失范行为靶向引导技术框架模型构建的原则

面向网络舆情靶向引导的技术框架模型的构建，是按照网络舆情信息受众、网络舆情事件与网络舆情信息受众失范行为三个核心概念组成的统计集合体。在整个框架的构建过程中，需要以数据存储智能化、模型衔接紧密化、运行流程集约化为构建原则。

数据存储智能化指在整个框架结构中，数据通过抓取后的优化存储问题。在数据获取后，大部分数据属于原始数据，经过数据清理与语义识别的过程转化为可用数据，而后再通过网络舆情信息受众失范行为画像与网络舆情事件感知，形成统一的标签体系。这些标签数据需要以表的形式存储在数据库中。同时网络舆情信息受众失范行为靶向引导策略也需要以标签数据的形式存储在数据库中。该数据库包括受众画像库、事件感知库与靶向引导策略库 3 个数据库。数据库之间具有互联互通和数据互操作的功能。

模型衔接紧密化指在整个框架中，包括网络舆情信息受众失范行为拟合与网络舆情信息受众靶向引导，存在多种数理模型，例如数据获取模型、语义识别模型、行为预测模型、事件匹配引导模型、受众分型引导模型、失范行为匹配引导模型等，各模型之间需要通过网络舆情靶向引导的技术框架模型进行衔接，形成统一的数据接口，达到从前序模型的输出到后续模型的输入的无缝衔接。

运行流程集约化指在整个框架中，模块的运行存在先后顺序的情况，也存在并行处理的情况，哪些模块可以并行处理，哪些模块必须遵从先后顺序，从网络舆情信息受众失范行为拟合与靶向引导的数据输入到数据输出要经过哪些步骤等。整体流程的合理化，不仅能够理清框架模型运行的流程，而且能够通过数据、模型的并行计算与处理提高框架的运行效率。

4 网络舆情受众失范行为靶向引导的技术框架

网络舆情受众失范行为靶向引导的技术框架模型是根据网络舆情受众、网络舆情事件与网络舆情受众失范行为三维体系构建而成，其中包含了数据获取模块、靶向引导策略库、网络舆情信息受众失范行为拟合模块、受众画像库、事件感知库、个体分型引导模块、行为匹配引导模块、事件匹配引导模块、触发点发现模块等要素。

框架的流程是从数据获取到行为拟合再到靶向引导策略输出等各模块相互衔接的运行过程。网络舆情信息受众失范行为靶向引导框架结构见图 1。下面分别从网络舆情信息受众失范行为拟合与靶向引导的框架模型要素及流程进行分析。

4.1 网络舆情受众失范行为靶向引导技术框架的模块构成分析

网络舆情受众失范行为靶向引导技术框架的模块构成包含了 7 个部分，分别是数据获取模块、失范行为拟合模块、数据库模块、个体分型引导模块、行为匹配引导模块、事件匹配引导模块、触发点发现模块等。具体模块的位置见图 1。

（1）数据获取模块。数据获取模块包括受众数据获取与事件数据获取两个部分。其中，受众发现模块，是对网络舆情信息受众的主页、发文、评论等信息进行摘取、收集与处理的过程；舆情监控模块是对网络舆情信息受众参与的网络舆情事件的博文、评论信息摘取、收集与处

理的过程；数据获取模块是整个网络舆情信息受众失范行为拟合与靶向引导框架的信息输入单元，通过网络爬虫工具进行实施，是无间断处理的过程。

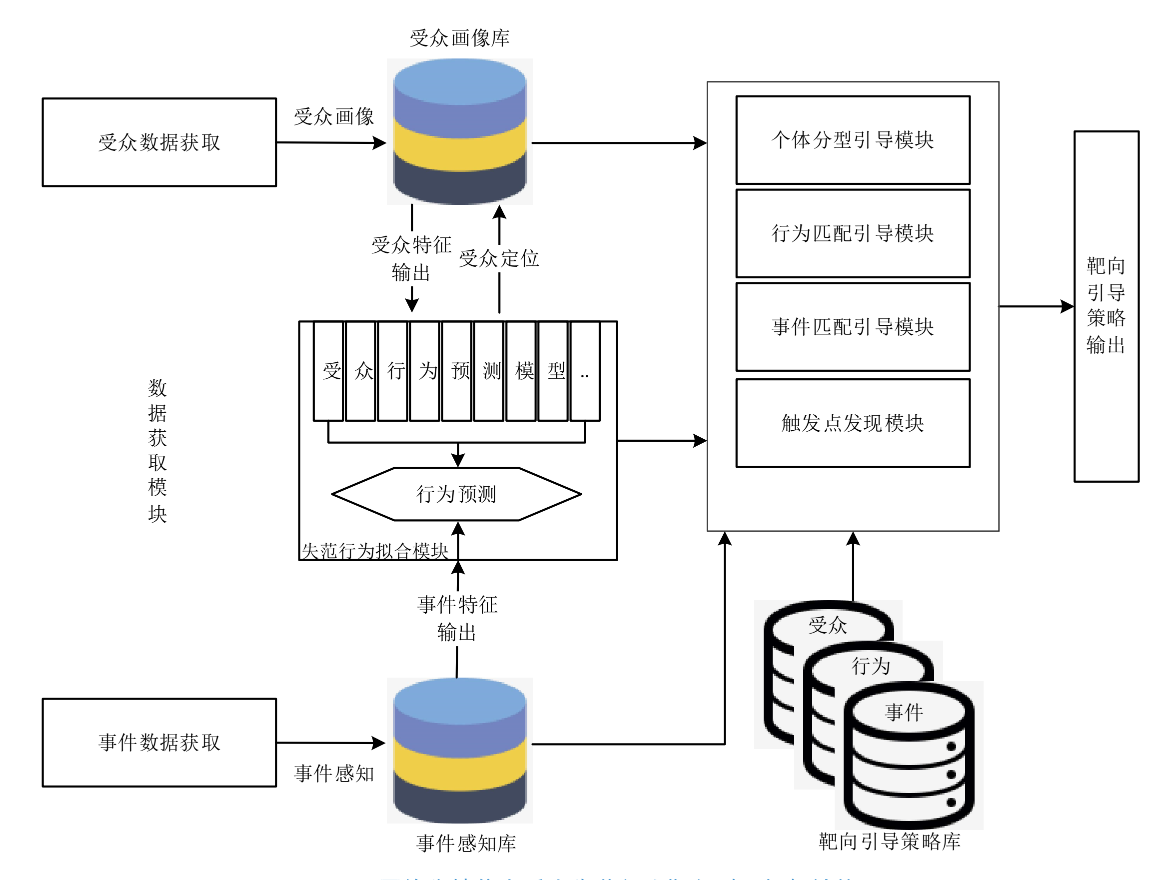


图 1　网络舆情信息受众失范行为靶向引导框架结构

（2）失范行为拟合模块。该模块包含网络舆情信息受众画像、网络舆情事件情境感知与网络舆情信息受众失范行为拟合 3 个部分，分别对应了受众画像模型、事件情境感知模型与受众失范行为预测模型。其中画像与情境感知的作用是针对网络舆情信息受众与网络舆情事件的二次处理数据进行分型匹配，并将加工后的数据存入数据库；网络舆情信息受众失范行为预测的作用是结合改进的 S-O-R(刺激 - 机体 - 反应) 模型，将网络舆情受众画像库中的多类受众作为机体群，令每一类机体分别被事件情境刺激，对其发生的行为进行预测，并发现失范行为的过程；网络舆情信息受众失范行为拟合模块是数据处理单元，为网络舆情信息受众失范行为靶向引导做出铺垫。

（3）数据库模块。数据库模块包含受众画像库、事件感知库与靶向引导策略库。其中受众画像库与事件感知库建立在网络舆情信息受众画像与网络舆情事件情境感知的基础之上，在对应的标签体系下构建而成的用户分类集合与事件分类集合；靶向引导策略库是靶向引导策略的集合，包含了引导策略集合与引导路径集合；引导策略集合与引导路径集合均是按照受众、行为与事件三大要素进行标注。针对不同的受众、不同的行为与不同的事件均有不同的特定引导策略与引导路径；数据库模块是面向网络舆情靶向引导的技术框架模型的数据存储单元，是实施网络舆情信息受众失范行为靶

知识管理论坛，2020(3):159-174
DOI：10.13266/j.issn.2095-5472.2020.015

向引导的数据基础。

（4）个体分型引导模块。个体分型引导模块是连接网络舆情信息受众画像与网络舆情受众失范行为靶向引导策略的接口模块，具有针对网络舆情信息受众个体的独有特征，从靶向引导策略库中自动匹配引导策略与引导路径的功能。

（5）行为匹配引导模块。行为匹配引导模块是连接网络舆情信息受众失范行为拟合与网络舆情受众失范行为靶向引导策略的接口模块，具有针对不同的网络舆情信息受众失范行为从靶向引导策略库中自动匹配引导策略的功能。

（6）事件匹配引导模块。事件匹配引导模块是连接网络舆情事件情境感知与网络舆情受众失范行为靶向引导策略的接口模块，具有针对不同的网络舆情事件从靶向引导策略库中自动匹配引导策略的功能。

（7）触发点发现模块。触发点发现模块是实施靶向引导的启动器，不同的网络舆情事件与不同的网络舆情受众所启动靶向引导的条件也不尽相同，根据网络舆情信息受众失范行为预测拟合的结果，自动计算触发条件。

4.2 面向网络舆情受众失范行为靶向引导技术框架的运行流程分析

网络舆情信息受众失范行为拟合与靶向引导流程的基本步骤是数据更新、行为拟合、受众定位、策略匹配、引导触发与靶向引导6步。

（1）数据更新。数据更新即受众数据获取与事件数据获取两部分不断收集与更新数据的过程。数据更新的过程以抓取模型为基础，按照从热点舆情事件内容抓取到热点舆情事件受众抓取的方式展开。数据更新的过程是按照时间进程不间断进行。由于微博具有反爬虫机制，会进行封号操作，因此数据抓取的频率略微降低，并且每隔一个时间间隔进行下一次事件抓取。数据更新后，网络舆情事件数据通过情境感知模型进入情境感知库。在情境感知库中，若存在该舆情事件，则对该事件进行数据更新；若不存在，则添加事件并初始化。网络舆情信息受众数据通过受众画像存入受众画像库。在受众画像库中若已存在该受众，则进行数据更新；若不存在该受众则添加受众并初始化。

（2）行为拟合。通过数据获取模块的不断执行，新的网络舆情信息受众不断通过受众画像存入受众画像库；新的舆情事件也不断通过情境感知进入事件感知库。与网络舆情信息受众进入受众画像库不同，新的舆情事件经过筛选之后，危机等级高的事件会作为S-O-R模型中的刺激部分，对画像库中的对应网络舆情信息受众产生影响，调用合适的网络舆情信息受众失范行为预测模型对受众可能发生的行为进行预测，形成网络舆情信息受众失范行为拟合结果。行为拟合的结果表现为输出事件信息与行为预测信息。

（3）受众定位。根据网络舆情信息受众失范行为拟合结果，发现可能产生极端行为或者造成恶劣影响行为的网络舆情信息受众个体，在受众画像库中进行受众定位，输出受众信息。在受众定位的基础上，将受众与策略库中的策略进行匹配，有针对性地进行靶向策略引导。不同受众的行为发生时间与规律也不同，同时针对不同事件，触发引导与否或者触发引导的时机也不同，因此针对特定受众判断最佳的靶向引导触发时间也极为重要。

（4）策略匹配。利用网络舆情信息受众失范行为靶向引导模型，针对之前输出的受众信息、网络舆情事件信息与行为预测信息在靶向引导策略库中进行策略匹配，将匹配结果整合并准备输出。策略匹配的目的在于靶向引导，是对网络舆情信息受众特征、网络舆情事件特征与网络舆情信息受众失范行为特征进一步挖掘而产生的具有靶向性的机制。能够在最佳时机，以最准确的策略对事件、对受众进而对受众失范行为进行引导与管控。

（5）引导触发。在策略匹配与策略准备输出的过程中，利用触发点发现模型计算合理的引导触发时机，当时机达标后触发引导策略输出与引导路径输出。策略输出是对网络舆情事

件、网络舆情信息受众与网络舆情信息受众失范行为全局把控的前提下进行的，因此针对不同的情况，靶向引导需要考虑与准确判断触发与触发时间的条件。这样才能够精准地实施引导，避免盲目引导、引导不及时、引导过激、引导不准等问题的发生。

（6）靶向引导。将靶向引导策略与靶向引导路径输出给网络舆情控制专业人员提供指导。通过靶向引导达到适时、适当、适度、精准地完成对网络舆情信息受众失范行为的引导目的。适时指能够在舆情信息受众失范行为发生之前，采取措施引导受众发声；适当指采取准确的引导措施，不盲目进行引导；适度指采取的引导措施需要得当，不能过激，也不能不痛不痒；精准是指引导要精确锁定受众个体，根据受众个体的具体特征实施靶向引导。

在框架模型的运行过程中，行为拟合中的受众画像与情境感知、策略匹配与引导触发属于并行运行模块，各模块分别由不同的控制单元控制，并行计算。其他步骤均有时间先后顺序，网络舆情信息受众失范行为靶向引导流程如图 2 所示：

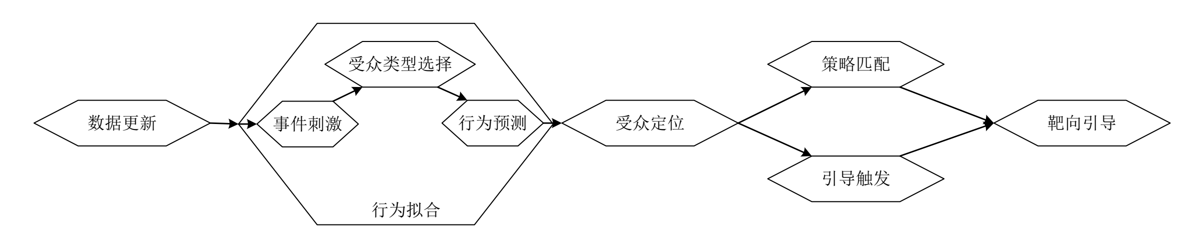


图 2 网络舆情信息受众失范行为靶向引导流程

5 面向网络舆情靶向引导的核心技术分析

5.1 数据处理技术

5.1.1 数据获取

首先，根据某事件为背景的所有微博进行筛选，选出事件转折点的典型微博，根据事件状态、当事人状态与微博评论量进行筛选，选出事件爆发阶段，当事人原创的评论最多、情感最为丰富的微博作为研究对象；其次，遍历该微博下的所有微博评论并进行抓取，在抓取的过程中，提取每个舆情信息受众的主页链接，采集舆情信息受众主页资料；再次，遍历舆情信息受众所发布的微博，并进行抓取；最后，在遍历的过程中，若发现其参与到其他网络舆情事件当中，则返回抓取的第一步，选取该事件为背景的所有微博进行筛选，由此形成一个环状的舆情信息抓取流程。

数据抓取应该以 S-O-R 理论为核心，仅抓取必要的数据进行研究。舆情事件与关键微博环节属于事件情境，是 S-O-R 理论中刺激的部分。受众资料与受众微博属于受众机体，是S-O-R理论中机体的部分。受众微博与微博评论属于机体反应，是 S-O-R 理论中的反应部分。整个微博数据抓取过程以舆情信息抓取流程环为框架，以 S-O-R 理论为核心，形成如图 3 所示的数据抓取模块。

总体来说，网络舆情信息数据抓取模块可以分为舆情事件、关键微博、微博评论、受众资料与受众微博等子模块，集中体现为刺激、机体与反应三部分。作为数据建模与数据分析的基础，有效的数据清洗、数据筛选与数据预处理等操作十分必要。

5.1.2 数据清洗

在数据抓取的过程中，由于数据来源问题、数据格式问题、数据抓取中的网络波动问题等，往往会造成抓取的数据形式混乱、格式不统一、数据重复、含有缺失值等情况 [14]。因此需要进行数据清洗的工作，使得数据完整并且具有统一的格式。

知识管理论坛，2020(3):159–174
DOI：10.13266/j.issn.2095–5472.2020.015

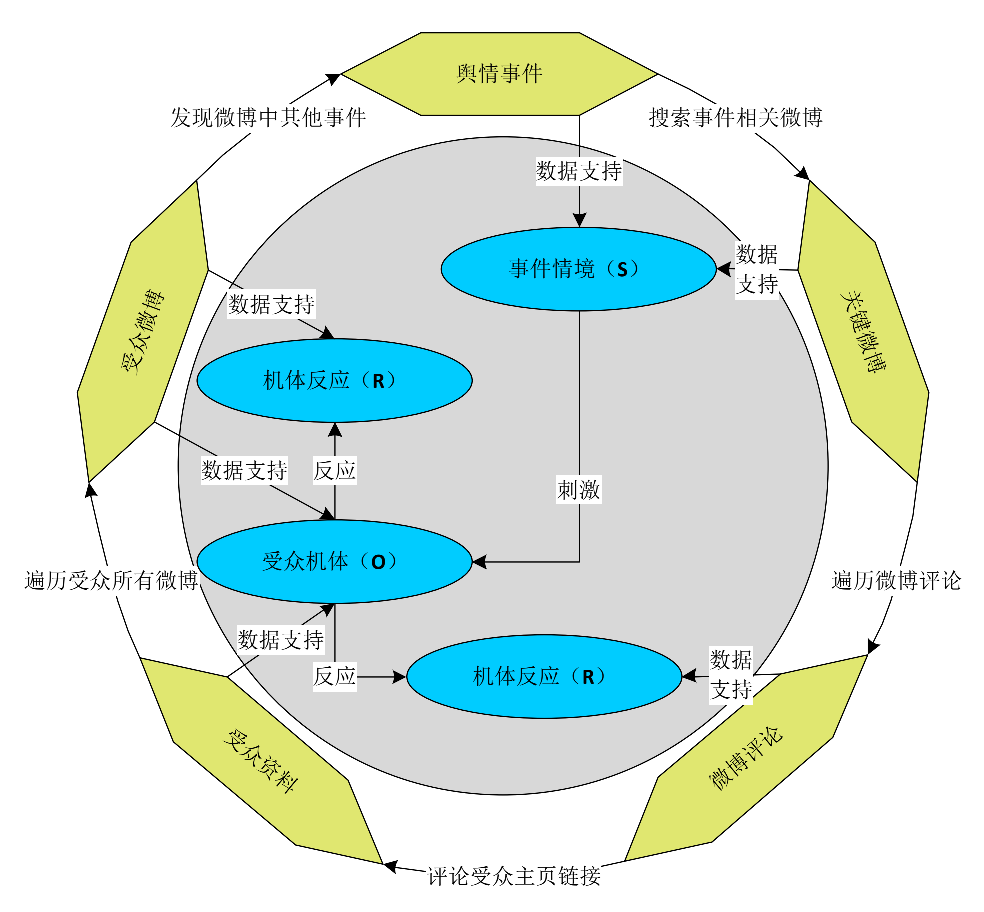


图 3 数据抓取模块示意图

笔者在数据清洗过程中，首先将舆情事件、关键微博、微博评论、受众资料与受众微博的不同抓取形式，分别以人为中心、以事件为中心与以文本为中心进行编号，形成人 - 文本 - 事件一一对应的数据集。

针对不同的数据集，将其中文本格式的部分统一为 UTF-8 格式；将含有数字的部分统一定义为 Double 格式；将含有时间的部分，统一按照 yyyy/mm/dd 的格式整理；将所有数据单元格应用 Strip 函数，去除多余的空格。

由于抓取中网络时延的问题与数据来源的问题，出现了大量缺失值。一般的缺失值处理办法有全局常量填充、中心度量填充、同组均值填充、最有可能值填充等。笔者针对缺失值问题，做如下处理：由于缺失值既含有文本数据也含有数字数据，因此统一将含有缺失值的行删除，方便后续处理。

5.1.3 分词

笔者采用 Python 编程技术，应用中文分词工具 jieba 分词，针对文本内容实现中文分词。考虑到如代词、介词、语气词等一系列对推文本身语义无关的特殊词语，笔者采用 jieba 中的停用词工具，定向分词。在此次实验中只采用分词结果中的一般名词、人名、音译人名、地名、音译地名、机构团体名、其他专名、名词性惯用语、名词性语素、新词、处所词、一般动词、副动词、动名词、动词性语素、形容词、副形词、名形词、副词、区别词这 20 类词。并添加“有”“没有” “还” “是” “也”等未滤出的停用词 [15]。

5.1.4 词向量

在分词的基础上，将人类可以理解的文本信息，转化为计算机能够理解的信息，成为



自然语言处理中的一项关键技术。最初的词向量模型是将不同词语的存在与否以（1,0）的方式展示在矩阵中，但是以这种方式形成的词向量矩阵过于稀疏，并且缺乏字词之间的相似关系描述。因此 2013 年 T. Mikolov 等提出了 Word2vec 模型 [16] 可以快速而高效地训练词向量，体现词与词之间的关联度关系。Word2vec 模型包含两种基于神经网络的训练模型，一是 CBOW（Continue bag of word）模型，另一种是 Skip-gram 模型。CBOW 模型是通过上下文来预测当前词，而 Skip-gram 模型则是通过当前词来预测其上下文。两种相反的训练方法对应了不同的数据需求。CBOW 在小型语料库中表现良好，而 Skip-gram 则在大型语料库表现更为出色。由于两种算法在针对大量数据的时候，参数训练的规模都空前的巨大，极其耗费时间。因此 T. Mokolov 引入了两种优化算法 Hierarchical Softmax 和 Negative Sampling。两种训练算法与两种优化算法相结合可以得到 4 种框架。

笔者采用腾讯 AI 实验室训练的 200 维词向量模型 [17]。该模型是 Skip-gram 模型的优化版本，称为 Directional skip-gram。Directional skip-gram 算法是在文本窗口中词对的共现关系基础上，额外考虑词对的相对位置，从而提高词向量语义表示的准确性，具有训练样本丰富、语料新鲜、准确度高的特点。但是应用大规模数据训练，包含体量巨大的关键词序列和 200 维的词向量表示，在提升准确率情况下，无法满足应用词向量过程的提高效率问题，使得运用词向量模型进行语义识别的运算速度缓慢。而应用神经网络模型只需要进行一次模型的训练即可得到神经网络的训练参数，之后进行语义识别的过程中，只需要调用训练参数即可，并不会影响识别的效率。

5.2 行为拟合与匹配技术

5.2.1 基于多尺度卷积神经网络语义识别算法

网络舆情信息受众语义识别模型从宏观上来说，是利用卷积神经网络模型，将网络舆情信息受众发布的中文语料进行分类的过程。从微观上讲，是将网络舆情信息受众发布的评论、原创微博通过词向量处理封装成计算机可读的数字化信息，之后利用卷积神经网络的语义识别与分类功能将其按照情感强度分为：特低、低、中、高、特高 5 个等级；按照情感极性分为消极、中立与积极 3 个类别；按照话题类别分为：社会民生、政治经济、文化娱乐、教育科技、自然灾害、意识形态、公共安全、司法警务与其他。由于自然语言在经过词向量模型加工后，词向量之间的距离具有远近关系，因此分析一句话的语义不单单要考虑单个词语的含义，还需要将多个词语联系起来进行综合考虑。卷积神经网络中的卷积核具有按步长扫描的作用，可以通过控制卷积核的大小达到多个词语综合分析的效果。本文在 X. Yang 与 K. Shrivastava 提出的卷积神经网络模型 [18-19] 的基础上，利用不同的卷积核尺寸，构建了多尺度卷积神经网络模型，如图 4 所示：

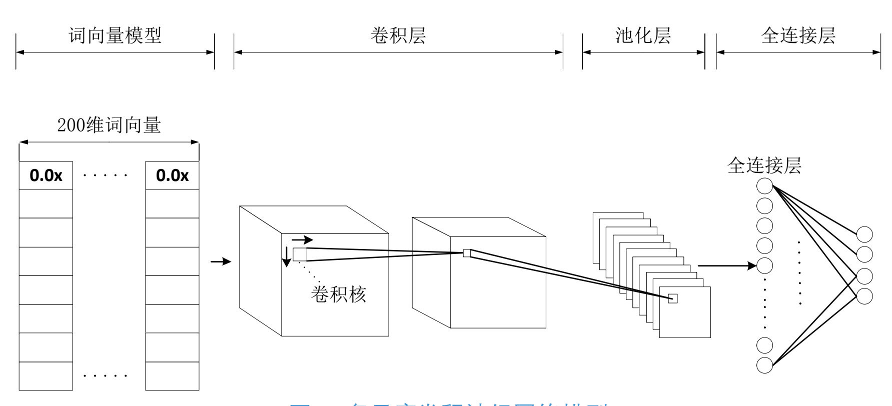


图 4 多尺度卷积神经网络模型

知识管理论坛，2020(3):159–174
DOI：10.13266/j.issn.2095–5472.2020.015

5.2.2 基于用户画像与情境感知模型的网络舆情受众建模与事件建模

用户画像与情境感知实质上具有相似性，用户画像即对网络舆情受众个体的特征进行结构化的标签描述，解析该舆情受众的基本特征、行为特点、爱好偏好等。情境感知即对网络舆情受众所处的环境进行解析，实质上是对网络舆情受众所面对的舆情事件进行标签化描述，分析事件的类别、属性与其对舆情受众的影响等。用户画像与情境感知离不开语义识别算法，在网络上无论是网络舆情受众或是网络舆情事件，均是由文字、图像、视频等异构多媒体数据组成 [20]，这就需要特定的语义识别算法对这些数据进行解读。笔者研究以文字数据处理为主，应用前一节提及的语义识别模型来进行处理。用户画像与情境感知建模如图 5 所示：

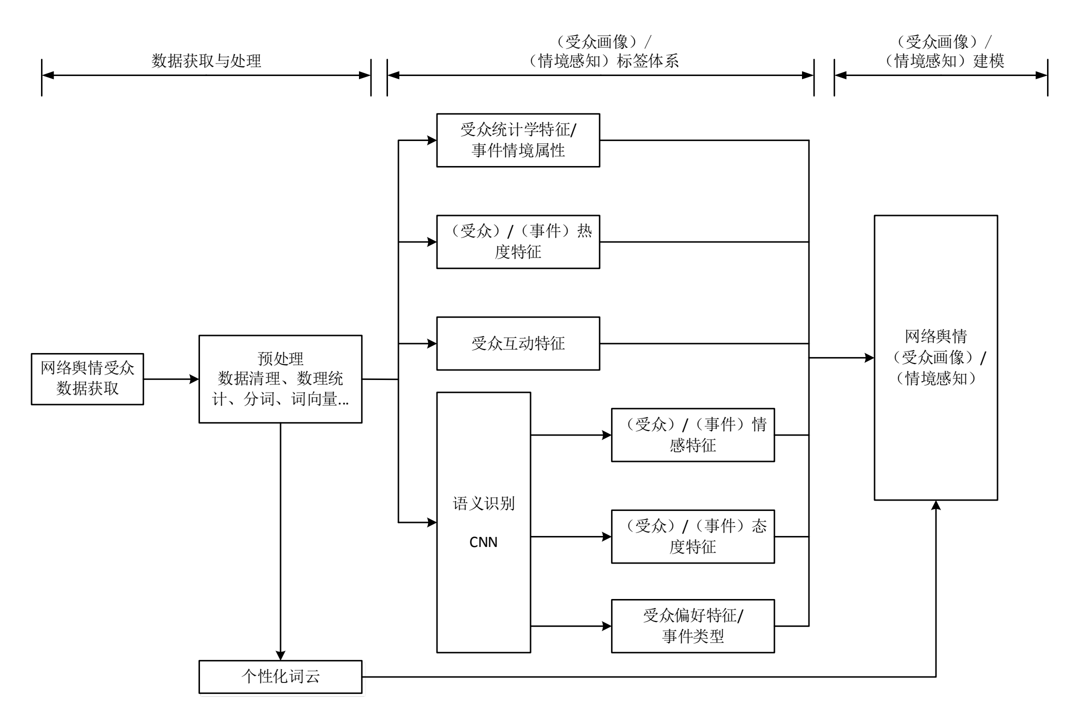


图 5 用户画像与情境感知建模

5.2.3 基于贝叶斯网络的网络舆情受众信息行为预测算法

贝叶斯网络是基于贝叶斯定理的一类概率模型，可以用来解决分类问题，它能够用图像的方式体现出数据之间的相互关系并且能够给予事件概率化的描述。在网络舆情受众信息行为预测的研究当中，利用贝叶斯网络可以将网络舆情信息受众与网络舆情事件结合在一起进行分析，即利用前一节提及的用户画像模型与情境感知模型，得出具体网络舆情信息受众在特定网络舆情事件下表现出情感强度、情感极性、行为偏好、行为表现的类型，从而计算出该网络舆情受众产生特定失范行为的概率。具体贝叶斯网络模型如图 6 所示：

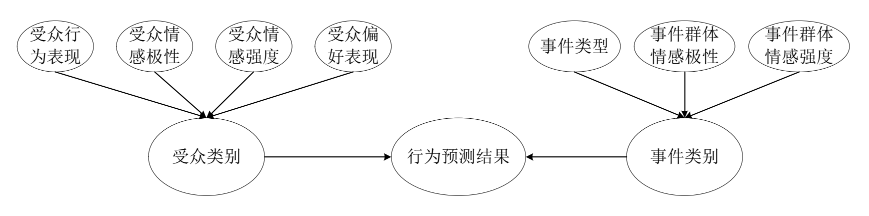


图 6 基于贝叶斯网络的网络舆情受众信息行为预测模型

5.3 数据库技术

网络舆情受众失范行为靶向引导的技术框架模型中，包含 3 种类别的数据库，分别是受众画像库、事件感知库与靶向引导策略库。其中受众画像库与事件感知库，均是对具体网络舆情受众与事件的标签化描述。靶向引导策略库是按照网络舆情受众、网络舆情事件与网络舆情受众失范行为的历史记录提出的引导策略集合。

数据库的设计按照从抽象到具体可以分为数据库的概念设计、数据库的逻辑设计与数据库的物理结构设计 [21]。下面以网络舆情受众画像库为例简述数据库的概念设计、逻辑设计与物理设计。

5.3.1 数据库概念设计

数据库的概念设计是应对数据库需求的概念设计模式。在数据库的概念设计中，需要在数据库的需求分析基础上，构建数据库概念模型，即 E-R 图。在网络舆情受众失范行为靶向引导技术框架模型中，需要设计 3 类数据库，笔者将在微博平台上，以网络舆情受众画像库为例，简述数据库概念设计。

在网络舆情受众画像库中，需要整合网络舆情受众的历史记录，包括受众主页、受众评论、受众微博与网络舆情受众。其中网络舆情受众与受众主页是 1 对 1 对应的实体，受众通过编辑主页，可以自主修改受众昵称、所在地、性别等基本信息，这些信息均通过受众主页抓取获得。网络舆情受众通过发布行为，如发布或者转发微博，网络舆情受众与微博是 1 对 n 对应的实体，一个网络舆情受众会发布或转发多个微博，而微博包含了微博 ID、微博内容、转发数等数据，这些数据来源于数据抓取模型。网络舆情受众通过评论行为，对特定微博进行评论，评论与网络舆情受众存在 n 对 1 的实体对应关系，一个网络舆情受众会发布多条评论信息，评论信息的采集来源于微博评论信息采集过程。微博与其评论也存在 1 对 n 的实体对应关系，一条微博往往包含多条评论。通过上述概念分析，可以得出如图 7 所示的网络舆情受众画像数据库概念设计 E-R 图：

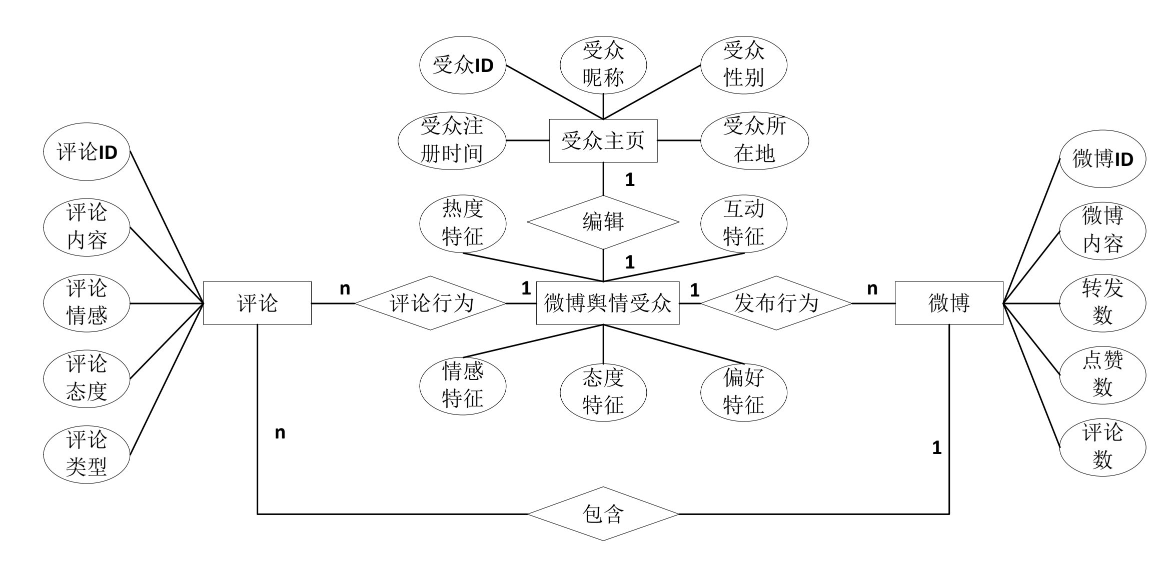


图 7 网络舆情受众画像库数据库概念设计 E–R 图

5.3.2 数据库逻辑设计

数据库的逻辑设计，是将概念关系转变为逻辑关系的过程。数据库的逻辑设计过程中，要满足 3 种范式，才能达到消除数据冗余、提高数据库效率的作用。

第一范式：数据库表是字段只含有基本数据类型的单一属性的二维表。即数据库表中的所有字段，不能含有多层含义与多个属性，并

知识管理论坛，2020(3):159–174
DOI：10.13266/j.issn.2095–5472.2020.015

且每个字段均能够通过 CHAR、VARCHAR、INT 等基础属性表示。

第二范式：每个表只存在一个主键。任何一个数据库表的主键，不能重复，只含有唯一值。

第三范式：消除对主键的传递依赖。例如 C 列与 B 列具有依赖关系，B 列又与 A 列具有依赖关系的情况，需要分裂成多个表进行表述，不能存在于同一个表。

以网络舆情受众画像库中的舆情受众与微博的对应关系为例，通过表 1 与表 2 简述数据库的逻辑设计：

表 1 舆情受众示意表

实体	属性	描述	数据类型	可否为空	键
舆情受众	受众ID	舆情受众唯一编号	BIGINT	FALSE	主键
	月均关注数	描述舆情受众每月关注次数	DOUBLE	FALSE	
	月均微博数	描述舆情受众每月微博发布次数	DOUBLE	FALSE	
	月均点赞数	描述舆情受众每月点赞次数	DOUBLE	FALSE	
	月均评论数	描述舆情受众每月评论次数	DOUBLE	FALSE	
	月均转发数	描述舆情受众每月转发次数	DOUBLE	FALSE	
	……				

表 2 微博示意表

实体	属性	描述	数据类型	可否为空	键
微博	微博ID	舆情受众唯一编号	BIGINT	FALSE	主键
	微博内容	描述舆情受众每月关注次数	VARCHAR	TRUE	
	转发数	该微博的转发数	BIGINT	FALSE	
	点赞数	该微博点赞数	BIGINT	FALSE	
	评论数	该微博的评论数	BIGINT	FALSE	
	作者ID	微博发布者ID	BIGINT	FALSE	外键
	……				

5.3.3 数据库物理设计

在数据库物理设计的过程中，应用 DDL（Data definition language，数据定义语言），在 MySQL 数据库中，进行创建数据库架构、创建表格框架、修改表格、删除表格等操作。之后通过 DML（Data manipulation language，数据操作语言），进行添加数据、修改数据、查询数据等操作。

5.4 面向网络舆情靶向引导的核心技术评价

网络舆情靶向引导的核心技术包含了数据处理技术、行为拟合与匹配技术和数据库技术。微博舆情数据处理技术包含数据获取、数据清洗、分词与词向量四部分。行为拟合与匹配技术包含语义识别、用户画像与情境感知、行为预测三部分。

数据处理技术最关键的部分为数据获取技术。数据获取技术的核心是网络信息雷达。网络信息雷达具有深度挖掘、持续采集、实时更新、远程更新、优化任务调度的性能，能够达到数据垂直检索与采集定向性、立体化、全面化和

准确化。

行为拟合与匹配技术最关键的部分为语义识别技术。语义识别技术的核心是卷积神经网络模型。卷积神经网络模型模拟了人类阅读文本信息的过程，实现了对文本信息语义的识别与理解。卷积神经网络具有自组织、自适应、自学习的特点，能够达到对语义的定向与精准的识别。

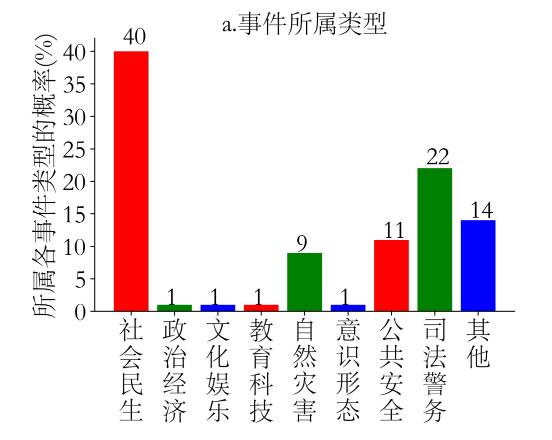
数据库技术是任何信息系统都不可或缺的重要组成部分。数据库的应用，保证受众、事件与策略的有效存储与有效调用。数据库设计的过程中，应用范式，最大限度地降低数据冗余，控制数据库规模，提高数据库运行效率。

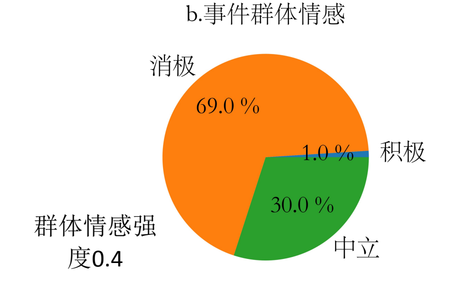
6 案例分析及框架验证

笔者以“李心草溺亡”事件为背景，对数据与模块不进行具体描述，仅依照框架流程进行验证。

6.1 数据更新

在数据更新的过程中，通过数据获取模块，获得“李心草溺亡”事件的微博事件数据与网络舆情受众数据，其中情境感知结果见图 8，受众画像结果见图 9。根据数据样本，将其存入情境感知库与受众画像库。从图 8 可以得出，“李心草溺亡”事件属于社会民生类舆情事件，消极情感占据大多部分且情感强度较强，词频统计中“李心草”“溺亡”“警方”“大学生”等词语较为突出。从图 9 可以得出，具体网络舆情受众“mini 宝贝在一起”的行为偏好为社会民生类舆情事件，受众类型为普通网络舆情受众，并且该受众的原创发文数多于转发数，积极言论较多，情感强度较为温和，词频统计中“生活”“孩子”“新闻”等词语较为明显。





李心草溺亡事件



图 8 情境感知结果示意图

知识管理论坛，2020(3):159-174
DOI：10.13266/j.issn.2095-5472.2020.015

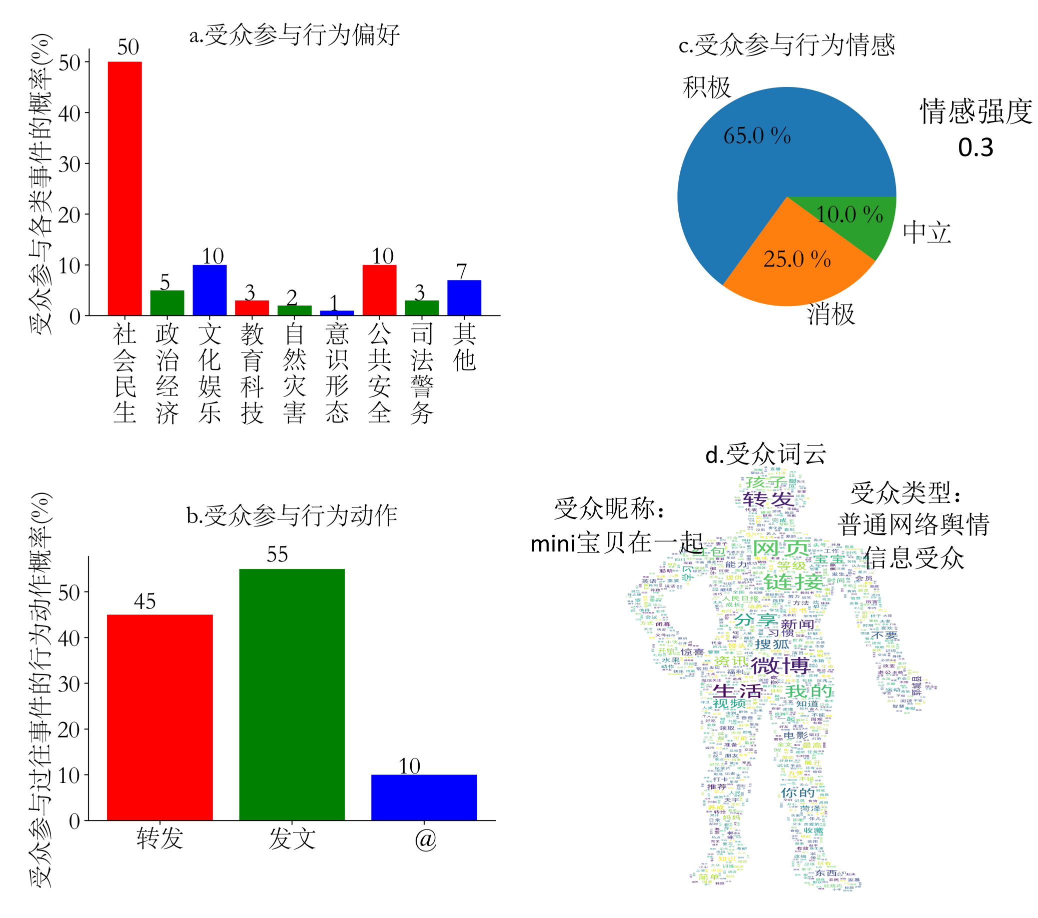


图 9 受众画像结果示意

6.2 行为拟合与受众定位

以“李心草溺亡”事件为分析对象，对受众画像库中的网络舆情受众进行拟合，可以得出如表 3 所示的行为拟合结果示意表。根据表中所描述的参与行为拟合结果，可以得出具体网络舆情受众与该事件的匹配度、参与行为的情感极性、参与行为的情感强度以及会对该事件进行发布、转发或者 @ 动作的可能性。根据匹配度可以对最可能参与此微博舆情事件的受众进行定位。

以受众“Runner- 游熙鹏”为例，可以得出该受众与“李心草溺亡”事件的匹配度最高，发布消极消息的可能性为70%，情感强度为0.28，采用发布动作的可能性为 72%。

6.3 策略匹配、引导触发与靶向引导

（1）事件策略匹配：根据事件策略匹配模块的计算结果，“李心草溺亡”事件，为中危、社会民生类微博舆情事件，系统结论为：应该对应中危微博舆情事件予以关注，并提出预警策略。预警策略包括以下 3 个方面：①跟踪预警策略，需要微博舆情管理者对其持续进行监控，按舆情发展情况考虑是否通知舆情应对机构；②协同引导策略，在舆情处理的过程中，通过与领袖型微博舆情信息受众及传统媒体型微博舆情信息受众进行合作，发挥其传播力强、影响范围广泛、可信度强的特点，使其为舆情朝积极有利方向发展进行发声；③持续关注策略，中危微博舆情事件危机程度不高，但是有可能演化为高危微博舆情事件，因此不能够掉以轻心，需要时刻对其进行关注，保持警惕。

表 3 行为拟合结果示意表

受众名	匹配度	积极	中立	消极	情感强度	发布	转发	@
Runner-游熙鹏	0.12	0.10	0.20	0.70	0.28	0.72	0.28	0.31
生命里的阳光明媚	0.10	0.00	0.15	0.85	0.31	0.49	0.51	0.02
阿唐的梦想在宁波	0.10	0.10	0.20	0.70	0.30	0.97	0.03	0.27
纷繁人生F	0.09	0.10	0.05	0.85	0.29	0.77	0.23	0.34
shishangshishang	0.08	0.05	0.10	0.85	0.35	0.52	0.48	0.21
廿三光景	0.07	0.20	0.05	0.75	0.32	0.66	0.34	0.27
邪毛-	0.06	0.05	0.20	0.75	0.32	0.84	0.16	0.37
清浅且薄	0.06	0.05	0.10	0.85	0.30	0.55	0.45	0.47
复读机机机机机	0.06	0.15	0.15	0.70	0.29	0.93	0.07	0.32
镜子看见自己	0.05	0.25	0.15	0.60	0.34	0.52	0.48	0.19
临村怪蜀黍	0.05	0.05	0.15	0.80	0.33	0.81	0.19	0.32
别浪费了我的药	0.05	0.05	0.15	0.80	0.33	0.97	0.03	0.07
问荆杨梅	0.05	0.05	0.10	0.85	0.34	0.64	0.36	0.10
我的大号不见了怎么办	0.05	0.00	0.10	0.90	0.31	0.97	0.03	0.08
在我目视的地方	0.05	0.10	0.20	0.70	0.33	0.90	0.10	0.30
虫倒虫不倒	0.05	0.15	0.25	0.60	0.36	0.90	0.10	0.56
鱼某人oo	0.05	0.20	0.10	0.70	0.33	0.70	0.30	0.30
胡歌老粉	0.05	0.15	0.20	0.65	0.34	0.85	0.15	0.14

（2）受众策略匹配：微博舆情信息受众“Runner- 游熙鹏”，是普通微博舆情信息受众。系统结论为：应该对应普通微博舆情信息受众予以关注，采用受众沟通策略。受众沟通策略包括以下 3 个方面：①加强教育策略，通过媒体宣传、媒体科普等方式，加强普通微博舆情信息受众的教育，提高普通微博舆情信息受众在面对微博舆情事件时的应对能力，减少其失范行为的发生；②靶向信息推送，在平时多向该受众推送具有积极意义的微博内容，为其创建和谐的微博环境；③切断其不良信息的获取渠道，尽量切断该微博舆情信息受众与不良受众之间的联系与信息获取渠道，从而营造和谐的微博环境，避免该受众失范行为的发生。

（3）行为策略匹配：“Runner- 游熙鹏”对于“李心草溺亡”事件，将要采用的参与行为属于情感强度较低的消极发布行为，属于“特低危失范行为”。系统结论认为：目前可以对其采取观察策略，并没有直接或者间接干预手段的必要。

7 总结与展望

笔者构建网络舆情受众失范行为靶向引导的技术框架，并且总结利用 3 种核心技术，使得提出的框架从技术上实现成为可能。通过以“李心草溺亡”事件为例，梳理了框架运行的流程，计算出框架运行的结果，并且针对运行结果提出了具体的靶向引导策略，验证了框架的有效性。但是，笔者提出的网络舆情受众失范行为靶向引导技术还处于研究初期，还有较多的内容没有融合从而没有形成更加完整的体系。在今后的研究中，笔者将从上述七大模块入手，构建并优化成型的数理模型，进行更加深入的研究。

知识管理论坛，2020(3):159–174
DOI：10.13266/j.issn.2095–5472.2020.015

参考文献：

[1] 曾润喜，王国华，徐晓林．高校网络舆情的控制与引导[J]. 情报理论与实践，2009, 32(11): 79-82.
[2] 吴红．高校图书馆如何利用流行文化积极元素引导大学生践行社会主义核心价值观[J]. 图书情报工作，2015, 59(S1): 143-146.
[3] XIA W. Occupy media highland, lead the consciousness development of college students[J]. Canadian social science, 2015, 11(2): 86-93.
[4] 郑磊，任雅丽．中国政府机构微博现状研究[J]. 图书情报工作，2012, 56(3): 13-17.
[5] 罗潇潇，何跃，熊涛．突发公共事件中权威信息对微博内容的影响研究——以柳州镉污染事件为例[J]. 图书情报工作，2012, 56(11): 123-127.
[6] 王晰巍，邢云菲，王楠阿雪，等．新媒体环境下突发事件网络舆情信息传播及实证研究——以新浪微博“南海仲裁案”话题为例[J]. 情报理论与实践，2017, 40(9): 1-7.
[7] 赵晓航．基于情感分析与主题分析的“后微博”时代突发事件政府信息公开研究——以新浪微博“天津爆炸”话题为例[J]. 图书情报工作，2016, 60(20): 104-111.
[8] 刘永，许烨婧．面向情境的情报服务理论问题研究[J]. 情报理论与实践，2013, 36(11): 1-4, 19.
[9] 刘英杰，黄微，闫璐．基于A–V–P的网络舆情信息情感维度空间的模型构建[J]. 情报资料工作，2017(6): 12-18.
[10] 曾胜泉．网络舆情应对技巧[M]. 广州：广东人民出版社，2019:100-241.
[11] 鱼震海．基于新媒体环境下网络媒体失范行为的分析研究[J]. 现代情报，2013, 33(8): 172-174, 177.
[12] 谢晖．我国网络社会失范行为的法律规制——以网络舆情为视角[J]. 法制博览，2019(31): 37-39.
[13] 杨祺．微博意见领袖的行为失范及对策研究[D]. 长沙：湖南大学，2016.
[14] 黄微，许烨婧，刘熠．大数据环境下多媒体网络舆情并发获取的数据驱动机理研究[J]. 情报理论与实践，2019, 42(6): 42-48, 16.
[15] 黄微，刘熠，许烨婧，等．网络舆情推文的热度测度模型构建[J]. 图书情报工作，2019, 63(20): 17-25.
[16] MIKOLOV T, CHEN K, CORRADO G, et al. Efficient estimation of word representations in vector space[EB/OL]. [2018-06-17]. https://arxiv.org/pdf/1301.3781v3.pdf.
[17] YAN S, SHUMING S, JING L, et al. Directional skip-gram: explicitly distinguishing left and right context for word embeddings[C]//The 2018 Conference of the North American Chapter of the Association for Computational Linguistics. New Orleans: Association for Computational Linguistics, 2018: 175-180.
[18] YANG X, XU S, WU H, et al. Sentiment analysis of Weibo comment texts based on extended vocabulary and convolutional neural network[C]// 2018 international conference on identification, information and knowledge in the internet of things. Beijing: Procedia Computer Science, 2018: 361-368.
[19] SHRIVASTAVA K, KUMAR S, JAIN D K. An effective approach for emotion detection in multimedia text data using sequence based convolutional neural network[J]. Multimedia tools and applications, 2019, 78: 29607–29639.
[20] 黄微，刘熠，孙悦．多媒体网络舆情语义识别的关键技术分析[J]. 情报理论与实践，2019, 42(1): 134-140.
[21] 黄微，赵江元，许烨婧．多媒体网络舆情知识库群互操作要素及机理研究[J]. 图书情报工作，2019, 63(20): 43-50.

作者贡献说明：
黄　微：负责论文框架设计与内容指导；
刘　熠：负责论文撰写与模型搭建；
郭苏琳：负责论文修改与校对。



知识管理论坛
2020 年 第 3 期(总第 27 期)

Technology Framework Model of Target Guidance of Anomie Behavior of Network Public Opinion Audience

Huang Wei[1] Liu Yi[1] Guo Sulin[1,2]

[1]School of Management, Jilin University, Changchun 130022

[2]School of Architecture and Urban Planning, Jilin Jianzhu University, Changchun 130119

Abstract: [Purpose/significance] The anomie behavior of the audience is the crucial factor of drastic changes in public opinion. This paper aims to construct the technical framework of targeted guidance of anomie behavior, and guide the anomie behavior from the technical methods. **[Method/process]** This paper started with the analysis of the target guidance of the anomie behavior of the Internet public opinion audience. Then established the framework model and described each module of the framework model in detail. In addition, this paper organized the core technologies according to each module of the framework, and put forward the further research ideas. Finally, this paper verified the framework by taking microblog public opinion as an example. **[Result/conclusion]** Targeted guidance is a method combining theory and application. The technology framework model of target guidance for the anomie behavior of Internet public opinion audience can solve the specific defects of current guidance strategy from the technical level, and enhance the operability of targeted guidance.

Keywords: network public opinion   anomie behavior   target guidance